EOSDIS Technology Assessment Study

WAN Network Technology

TonyLisotta
Mark Radwin
NASA Ames Research Center

Objective

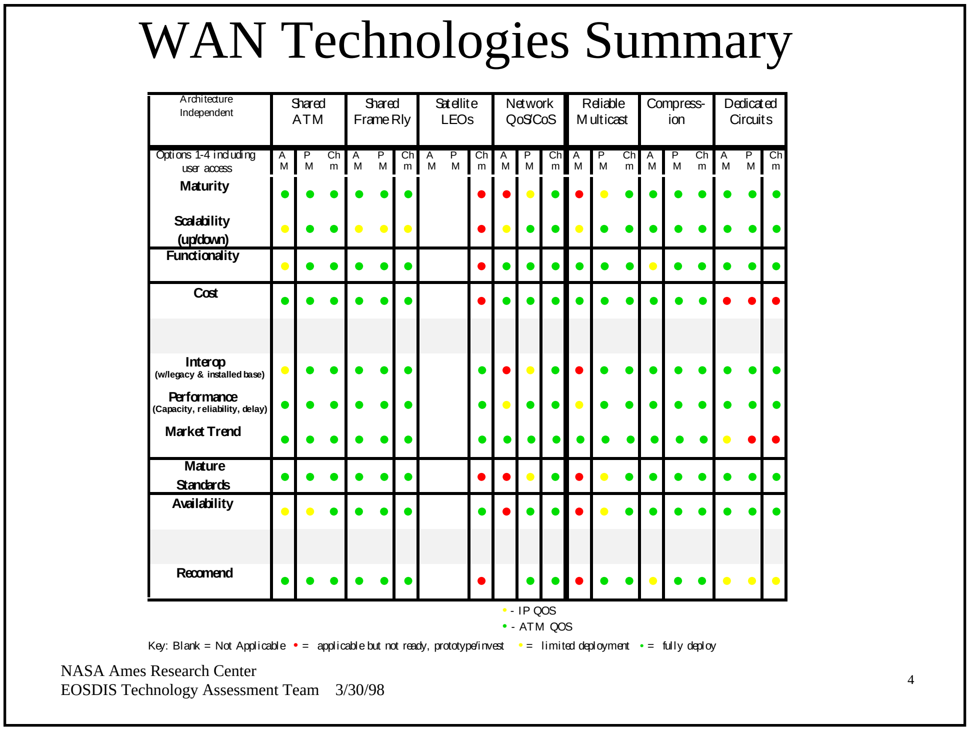
Evaluate and assess WAN technologies, as they apply to the architectures presented by EOS, against the criteria listed on the following page.

WAN Technologies Considered

- Shared ATM Networks
- Shared Frame Relay Networks
- Network QOS/COS
- Reliable Multicast
- Private/Dedicated Networks (T-carriers)
- LEO Broadband Data Satellites
- Compression

Assessment Criteria

- **Maturity** - Stable and reliable, wide vendor support,
- **Scalability** - Scalability with respect to the size of the network, and potential growth in performance
- **Functionality** - Does it work as defined by the standards, applicability to the EOSDIS architectures
- **Cost** - Relative dollar cost to implement and maintain the technology as well as cost required for training
- **Interoperability** - How well does the technology interoperate with legacy and installed base technologies
- **Performance** - The capacity, delay, QOS/COS & reliability of the technology
- **Market Trend** - Is the technology dominating, entering or leaving the overall network industry market
- **Mature Standards** - Are the core standards fully defined and accepted for the technology
- **Availability** - Can you buy this technology at or close to most critical sites



The data transmission volumes currently projected by the epochs of the launching of the major stream producing satellites, AM-1 in June, 1998, PM-1 in December, 2002, and AM-2 in June, 2004, will not of themselves pose challenges to present day WAN or LAN network transmission capacities. This can be easily seen by Table 1 below which depicts the total DAAC-to-DAAC data traffic requirements derived from the v4.0 (3/96) materials from the GSFC SPSO for all three major satellites (viz. post 2004 timeframe) and for all Level 1B and Level 2/3 data products. Note that these tables do not take into account the bandwidth savings over replicated unicast transmissions that will be provided by the single transmission possible through use of reliable multicast protocols as explained in more detail in a subsequent section.

	EDC	GSFC	JPL	LaRC	MSFC	NSIDC	Total
EDC		1402.537		19.291		3.477	1425.305
GFSC	52.511						52.511
JPL							0
LaRC		446.158					446.158
MSFC							0
NSIDC		21.915			5.1		27.015
Total	52.511	1870.61	0	19.291	5.1	3.477	1950.989

Table 1: Total DAAC-to-DAAC Data Traffic in 2004 in GB/Day

From the above table, the largest traffic generator/sink is GSFC at 1870.61 GB/day. This equates to 173.2Mbps assuming continuous transmission. Such a traffic load could easily be handled by two OC-3 links to NREN/NGI, before reductions obtained by lossless compression and reliable multicasting are taken into account. In one data pull experiment, we determined that ordinary and non-optimized Lempel-Ziv lossless compression yielded over a 4:1 compression ratio on the sample EOSDIS data files. This value is probably representative and might even be improved on by adjusting the Lempel-Ziv table sizes and compression block lengths or through use of other lossless compression technologies that are tuned to the specific data product structures. Assuming only the "off-the-shelf" Lempel-Ziv compression, the maximum traffic flow could be reduced to below 43.3Mbps which just exceeds the net capacity of a single T-3 45Mbps line, so that two T-3 circuits could comfortably handle the routine traffic and a single OC-3 connection could handle a 100-200% pull in addition to the baseline traffic.

Since the EOSDIS projectable traffic volumes per se won't themselves challenge networking technologies of the corresponding epoch, the rest of this networking technology review will address the important emerging technologies, their specific impacts on EOSDIS, where they should be utilized, and the potential for cost-savings and production time-compression they offer.

WAN Technologies

- **Shared ATM networks**:
  - mature if used as circuit replacement,
  - aggressive pricing by the carriers and highly scalability in capacity.
  - all functionality of ATM will not be available during the AM era- carriers won't be providing a production SVC offering...
  - scalability in terms of nodes on the network is an issue during the AM era. PVC based networks which are large and fully-meshed become hard to manage.
- **Network QoS**:
  - mature ATM level QoS, immature TCP/IP standards, and immature ATM-IP interop during the AM era
  - not widely installed until PM; recommend early adoption based on prototyping.
  - at that time it will provide mechanism for more efficient use of WAN bandwidth by allowing for prioritization of data products, over existing links (bandwidth sharing), to guarantee delivery.
  - ATM QoS *could* be used end-to-end *today* over ATM WANs (NREN/NISN, ESnet, vBNS, etc.) if campus backbone is ATM and uses either ATM-to-the-server/desktop.

WAN Technologies

- **Shared Frame Relay**:
  - in general a mature, cost effective WAN technology
  - not scaleable beyond T-3
  - frame relay to ATM functionality is still immature but is being used by production network currently
- **Reliable Multicast**:
  - still in R&D mode with some commercial protocols available
  - expect widespread maturity/availability by PM era
  - push in EOSDIS community by prototyping; supply technology to all subscribers when ready
  - significant savings of: internal network bandwidth, elapsed time to complete transmissions to all receivers, I/O and CPU at the server
- **Satellite LEOs**:
  - transmit a single copy of files to arbitrarily large number of receivers *simultaneously* rather than serially and repetitively thereby resulting in more efficient use of network bandwidth
  - global access to undeserved communities
  - unavailable until 2002; immature, costs unknown, etc.

WAN Technologies

- **Compression**:
  - seems to be a total winner from start
  - possibly establish compression/decompression processor pool out of low cost, fast, commodity PCs
  - save on WAN capacity requirements and therefore costs
  - many issues are raised when one talks about compression therefore it is recommended that EOS invest in benchmarking/prototyping efforts to establish the feasibility
- **Dedicated Circuits**:
  - minimize their use over long distances by maximizing use of government/public network connectivity and QoS for priority
  - very mature and very costly

Recommendations and Suggested Investment Opportunities

- AM era
  - Recommendations
    - migrate current dedicated services to ATM and Frame Relay as appropriate. Use ATM COS within the WAN network, and where ATM exists in the LAN, use it end-to-end.
  - Investment Opportunities
    - prototypes of reliable multicast and network QoS at both layers 2 and 3.
    - researching data compression for use in the EOS system
    - continue increasing intra-agency network bandwidth/cost sharing
- PM and Chem eras
  - Recommendations
    - continue to deploy ATM and Frame Relay within the WAN.
    - begin the production use of QoS/CoS and Reliable Multicast
    - implement compression within the system is applicable
  - Investment Opportunities
    - LEO Satellites

Where Research is being Done

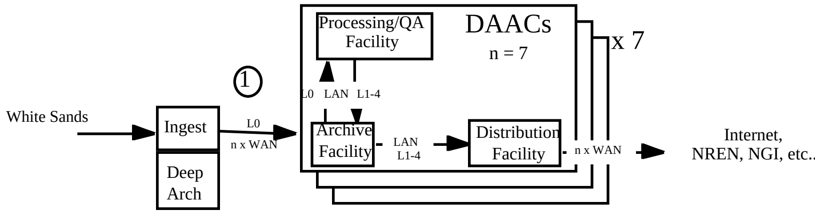
- QoS
  - DARPA has several funded QoS research locations
    - http://www.ito.darpa.mil/Research Areas 96/Quorum.html
  - NASA Ames
    - Information Power Grid (IPG), a component of the HPCC/CAS and IT/ACNS programs, is researching QoS as it pertains to the IPG project
    - NREN/NGI is chartered with researching QoS applications
- Reliable Multicast
  - NASA Ames
    - Code II is beginning a project to investigate and prototype RM, Radwin, et. al.
  - Lucent Technologies/Bell Labs e-cast middleware & RMTP protocol
    - probably the most robust protocol currently available; beta testing now
    - http://www.bell-labs.com/projects/e-cast/
  - Federal Multicast SIG Meeting (reliable multicast apps included)
    - http://www.starburstcom.com/rick/fedipmulticast.htm
    - NOAA Science Center, FAA, NASA Telemedicine - Chuck Doarn
  - StarBurst Communications - Multicast File Transfer Protocol (MFTP)
    - Financial Publishing Co., Promus Hotels, Toys 'R Us, General Motors
    - Partners: Cisco, Microsoft, Hughes Network Systems, AT&T Tridom, Platinum Technology
  - IP Multicast Initiative (IPMI) -- not reliable multicast; industry orgn.
    - http://www.ipmulticast.com/
    - NASA Ames Founding Member
    - 1997 IP Multicast Usage Report ($1475)
    - http://ipmulticast.com/collateral/sd_report_mailer.pdf

Architecture Considerations

General Architecture Considerations

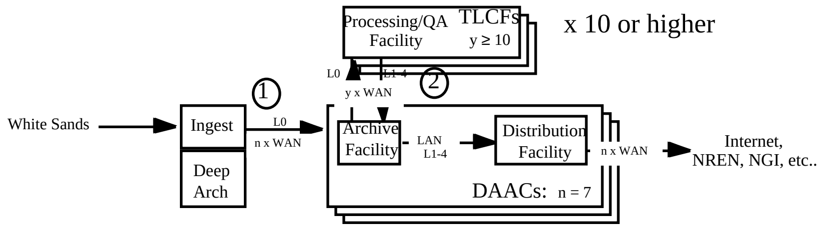
- For contingency reprocessing, use of overnight mailing of media may be a more cost effective solution than attempting to engineer enough capacity in the network to satisfy a reprocess and continue to produce current data products.
- For each architecture, WAN network efficiencies and cost saving can be achieved by locating as many DAACs and processing/QA and distribution centers as possible within in the same facility. e.g. in option 3 if a site is a DAAC and also has a TLCF, if they are in the same facility, only one WAN connection is needed.

Architectural Considerations - Opt 1



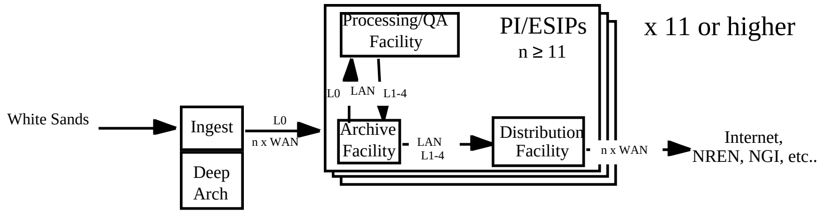
- 7 DAACs each with a WAN connection (e.g. NGI, NREN, etc.) to the Ingest Facility and to the Internet (e.g. NI, NREN) .
- WAN connection to Ingest facility must support the distribution of L0 data to the DAACs.
- Processing facilities at each DAAC need to communicate with other processing facilities at the other DAACs. This creates an (n * (n-1)) connectivity problem. For the example above the problem is , (7*6)=42 “unidirectional” connections are needed for direct processing-to-processing facility communication. This problem grows geometrically
- As a result of the amount of data being moved at point 1 above, a separate (virtual) private network is required, therefore the connectivity problem will need to be addressed. Should be easily handled by the PM-1 era with Public ATM network and reliable multicast of L0 and other multiply subscribed data products.
- Since each DAAC will archive L1-4 data products, query management across a WAN will have to be addressed.

Architectural Considerations - Opt 2



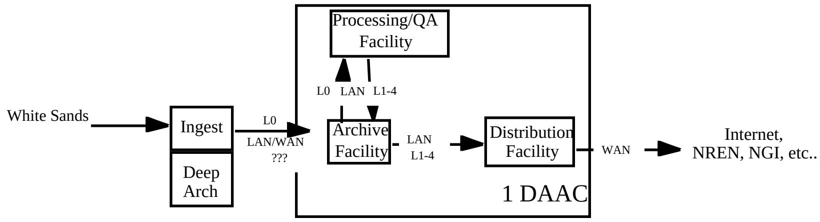
- 7 DAACs will need WAN connections (e.g. NGI, NREN, etc.) to the deep archiving facility, and the current 10 TLCFs will need WAN access to each of the 7 DAACs. Each DAAC will need Internet access for data distribution to users (e.g. NI, NREN)
- For the example above, the connectivity problem becomes [(n+y)*((n+y)-1)]; [17*16]= 272 “unidirectional” connections to connect all TLCFs and DAACs. This problem grows geometrically
- As a result of the amount of data being moved at points 1and 2 above, a separate (virtual) private network is required, therefore the connectivity problem will need to be addressed. Should be easily handled by the PM-1 era with Public ATM network and reliable multicast of L0 and other multiply subscribed data products.
- Since each DAAC will archive L1-4 data products, query management across a WAN will have to be addressed.

Architectural Considerations - Opt 3



- Each of the current 11 PI/ESIP location will need WAN connectivity to the deep archive and ingest facility as well as between one another. Internet access will be needed for distribution of the data products to users.
- Because each PI/ESIP location needs to be well connected, in the above example, the connectivity problem becomes, (n * (n-1)); 11*10=110 “unidirectional” connections to connect all PI/ESIPs. This problem grows geometrically.
- Query management across a WAN becomes a issue in this option if “one-stop” shopping is offered to the user community. This is due to each PI/ESIP maintaining an archive of their own data products; e.g. a user requesting a data product from the one-stop shop that requires the processing of several other data products, which reside at different locations (nested queries), could encounter a problem if the result of one of the nested queries is to large to transmit over the network, or if a network problem occurs.

Architectural Considerations - Opt 4



- The one DAAC only needs access to the Internet for the distribution of the data products.
- Amount of intersite production data moved across the WAN is considerably less in option 4 than in the other options. Pull impact is roughly the same.
- The challenge is to size the connections effectively to satisfy all transactions with regard to EOS data converging on this DAAC. Where in the other options, the user queries have the potential to be spread out over several archiving locations thereby reducing the aggregate amount of capacity needed at each site. Where in this option, ALL requests will be directed to this one DAAC.
- Assuring robust connectivity to major network peering points can mitigate this challenge
- Unclear if the ingest and deep archive center is located within the centralized DAAC or a WAN connection will be needed.

Supplementary Information

Shared ATM Networks

- Public/shared ATM services *aggressively priced*
  - bandwidth utilization can approach 100% - (“cell tax”+encapsulation)
- ATM with SONET provides a very scaleable solution for capacity
- ATM as circuit replacement is fairly mature, but has somewhat of a limited availability currently but will become more widespread over time
- Shared ATM network clouds: connect to closest Point Of Presence (POP), costs are *distance insensitive* from POP to POP and provide *fully meshed capacity between nodes*
  - ATM cloud services (Sprint’s NREN, ESnet)
  - vBNS/Internet2 connected sites will be internetworked to NREN
  - costs based solely on access line’s bandwidth and link to POP
  - assure long distance flows carried inside public/gov’t cloud services
  - access circuits are short and are hub to nearest wideband connected POP

NASA Ames Research Center
EOSDIS Technology Assessment Team 3/30/98 17

Asynchronous Transfer Mode (ATM) is a technology that has the potential to push the evolution of Local, Metropolitan and Wide Area networks beyond the current paradigm of looking at these environments as different, to developing a seamless environment that offers scaleable bandwidth, exceptional price points and the promise of being capable of providing end-to-end quality of service. However, as with any advanced protocol there is complexity that has to be dealt with in order to achieve the benefits that ATM can offer.

ATM is a connection-oriented cell switching technology that has fixed length cells defined. It is these characteristics together with the signaling protocols developed for ATM that allow it to be capable of providing the high-speed switching and quality of service. Furthermore, the cell-switching technology allows for the statistical multiplexing of may connection over the same physical media. This ability allows the ATM network service providers to make use of economies of scale and offer ATM services at prices equal to or better that a dedicated service, and with higher line capacities. Not only does ATM allow for higher capacity lines (primarily due to its use of SONET as an physical media) but it allows for tailoring these capacities to the exact requirements of the end-users. With dedicated circuits, on the other hand, one is faced with desecrate increments of capacity that can be purchased; the result being one usually pays for more than what is actually used.

In a inter- or intra-net environment, ATM is know as a non-broadcast multiple access network (NBMA) technology. With type of technology, an IP network designer has the flexibility to construct an IP architecture that best suits the needed of the users and does not necessarily have to conform to the physical constraints that other technologies such as dedicated circuits or FDDI rings force the designer deal with. This is most appealing in the WAN environment where a network service provider is provisioning the physical and link layer networks, and the customer is provisioning the network layer or IP layer. This is appealing because if one chooses to design a simple or flat IP architecture across the WAN, this adds complexity to the link-layer, which will be maintained by the network vendor who is more than likely staffed accordingly to be capable of dealing with this complexity. Furthermore, this technology suits an environment that is very dynamic. Because of the NBMA configuration and several standard guidelines produced by standards bodies, the addition, deletion or relocation of nodes becomes a simpler task because most of these tasks can be corrected by software changes, and require little physical reconfiguration.

Currently, ATM has been proven to be cost effective and reliable especially in the WAN environment. It should be noted that there is much debate about the viability of ATM, and this stems from the marketing promises of the early nineties that ATM will provide a network capable of supporting voice, video and data and have to ability to deliver QoS. While these promises are true, they have not all materialized fast as was hoped by the market. However, it has been shown that ATM, when use appropriately in the WAN, can achieve higher-available bandwidth than dedicated circuits at a much better cost.

Shared ATM Networks

- TCP/IP Virtual Private Networks (VPNs): emerging hybrid solution
  – contract w/ISP for capacity and service levels; secure (encrypted) tunneling through ISP's backbone; short connects to ISP's POPs
  – non-gov't connected sites can now use existing ISP path and bandwidth augmented with CoS/QoS to EOSDIS
- Leverage *existing* campus connectivity to Internet2, vBNS, ESnet, etc.
  – QoS/CoS will permit *prioritizing* EOSDIS production flows; other campus users continue to get *TCP/IP "best-effort" service*
  – economies of scale in bandwidth purchase, e.g. cost of 4 T-1 or 6Mbps capacity equals a T-3 or 45Mbps (NI rule-of-thumb) (8:1!)
  – ALL government agencies should *invest first in robust agency network interconnections & sharing site access connections*; last in duplicative site connectivity; EOSDIS will reap largest benefits
- Currently, DOE has a large ATM production network using Sprint

This acceptance of the technology can be shown by the growth in ATM switch capacity and port capacity by carrier switch manufactures. This growth is a direct result of the network communities buying and using these carriers ATM services. Another clear indication that ATM is a viable technology in the WAN can be seen in the government sponsored research testbeds. For example the current direction of the NREN/NGI programs are based on ATM WAN connectivity. These programs anticipate some connectivity at OC-48 rates in the 2000 time-frame and OC-3 to OC-12 connections to major science centers, universities (Internet2), and corporations in the nearer term. Both customer premise and carrier ATM switch capacities need to be characterized vs. time. Typical today is 10Gbps capacity, with 40Gbps easily anticipated in 1998.

On the other hand, ATM has not been a warmly accepted in the LAN desktop-connectivity environments because it is relatively more expensive than Ethernet and much less familiar. The unmatched success of switched 10Mbps and 100Mbps Ethernet has also limited the ATM-to-the-desktop deployment for at least several years to come. Once more applications begin to use APIs like WinSock v2 which supports native ATM end-to-end this may begin to change. Then a key factor will be the actual success of producing acceptable quality of service delivery in a variable-length packet (TCP/IP) environment.. ATM is gaining significant acceptance as a backbone technology in the LAN where it is replacing FDDI rings, and switched FDDI and switched 100Mbps Ethernet.

It will be challenged by Gigabit Ethernet starting in 1998, but will have a number of advantages and be a scaleable strategic solution in the high speed LAN/WAN environments. As ATM hardware vendors roll out the higher-capacity interfaces and switches, and the ATM QoS effort begins to emerge, and Multiple Protocol Over ATM (MPOA) cut-through routing becomes more widely installed, it is anticipated that ATM will then show a clear benefit to the LAN. One clear example of this advanced development of higher-capacity ATM switches can be seen at Stanford, where Prof. McKeown's Tiny Tera switch which is being funded by Cisco and will be using Texas Instruments CMOS technology. It is a coke-can sized switch with a maximum capacity of 32X32 ports with about 10Gbps/port, with an aggregate bandwidth of 200-500Gbps expected. Nortel and others have had terabit/second ATM switches in the laboratory for a few years. The next generation of "Big Fast Routers" will not be able to compete in capacity or price-performance, and it is likely that the two technologies will continue this divergence further into the future.

Quality of Service (QoS)

- QoS (to include class of service CoS) establish key parameters that prioritize traffic flows, establish bounds on such key parameters as delay, variation of delay, packet/cell loss rates, bandwidth available to flow, etc.
- To be effective, *QoS must be enforced end-to-end for entire path*
- Primary impact allows use of existing campus connectivity by prioritizing EOSDIS production flows -- possibly avoiding expansion of capacity, certainly avoiding expensive dedication of duplicative connectivity
- Applications need API support from ATM vendors or Microsoft's QoS Winsock v2 support for both TCP/IP and ATM, or equivalents
  - EOSDIS built/supplied software or SDKs should include ASAP

NASA Ames Research Center
EOSDIS Technology Assessment Team 3/30/98 19

Network Quality of Service (QoS) parameters or characteristics have a profound effect on determining the performance of network applications and services. There are somewhat different specific parameters and measures in each category of network technology that EOSDIS is likely to employ. The two principal network technologies of interest are TCP/IP and ATM based subnetworks. TCP/IP networks have traditionally provided a single quality of service, "best-efforts" delivery, reliability, and widely varying throughput, and delay depending on network engineering and the instantaneous offered loads. Control over the quality of service seen by applications is exercised by adequate provisioning of the network infrastructure. In contrast, a network with dynamically controllable quality of service allows individual application sessions to request network packet delivery characteristics according to their perceived needs, and may provide different qualities of service to different applications. New QoS protocols and protocol extensions are being defined in the Internet Engineering Task Force (IETF). The IETF's Integrated Services Internet (ISI) development activities are aimed at providing sufficient QoS controls so that bandwidth and delay-sensitive payloads like voice and video can be carried on the same network infrastructure as data. Integrated Services Internet developments worthy of mention are the ReSerVation Protocol (RSVP), Class of service Based Queuing (CBQ), Weighted Fair Queuing (WFQ), end-to-end QoS, and both aggregated and individual flow QoS contracts.The successful implementation of these types of services will go a long way toward eliminating the difficulties in EOSDIS production use of the non-differentiated service Internet as it exists today. Today's network technology as offered by ISPs could be used to design and engineer a Virtual Private Network (VPN) to handle the File Transfer Protocol (FTP) based model of delivering data products from a highly centralized group of processing centers within 1-week with a 98% probability of success. But not they are not suitable for more real-time, network-intensive, highly distributed architectures and data products with high data dependencies (e.g., use of precursor data products in producing additional ones) and tight production schedules, making heavy use of reliable multicasts, etc. This will benefit from the economics and capacity of the NREN, NGI and NI today, and will be further enhanced with the utilization of the network QoS technologies available today in ATM and their future implementations in TCP/IP ISIs in the next century.

There is an extremely powerful way in which the creation (in TCP/IP environments) and deployment of QoS facilities could easily and greatly reduce EOSDIS networking costs. It becomes very clear when one realizes that virtually all regular users of EOSDIS data products will already have some kind of internet connection for the myriad other applications that site/campus has for such connectivity. Currently, those connections if ATM access links (relatively rare today but increasingly common as more ISPs and LECs tariff them and additional public ATM network offerings come to market) already have per-virtual-circuit bandwidth specification/contract capability, but few router connected lines use the less fine-grained Class of Service (CoS) capabilities available today. For one thing, turning on those features has a significant impact on reducing the router's bandwidth capacity because they introduce a whole new (non-cacheable) category of packet-by-packet queuing manipulations.

With practical QoS features enable over these existing access links, it becomes possible to assign higher priority or guaranteed bandwidth and latency performance to the time-sensitive (e.g. reliably multicast data products traveling in either direction). If the percentage of total traffic into a site/campus due to EOSDIS subscription data products is low enough, there may be no need to upgrade the access line capacity. If there is need to add capacity, there are economies of scale that operate and the entire cost of access lines is not the burden of EOSDIS alone, but shared appropriately. This leaves the significant issue of QoS implementation in TCP/IP Internet services by the ISPs. They will certainly require additional compensation for these services. While the IETF has in recent years been attacking the protocol piece of QoS in TCP/IP, they have just formed a new working group to deal with the creation of "QoS servers" that will authenticate users and collect billing information. ATM has had the QoS (traffic management, class of service) features and cell-counting mechanisms built in from the beginning and will be able to support QoS in the WAN more easily and at a lower risk as a result. All of these QoS/CoS issues should be easily worked out and implemented by the EOSDIS CHEM-1 era and planning for and testbedding their implementation should begin immediately.

Quality of Service (QoS)

- TCP/IP QoS
  - IETF working on Integrated Services Internet technology specifications for TCP/IP networks; portions "completed", more work!
    - implementation requires *IETF completion & totally new routers and etherswitches* with built-in hardware for per-flow QoS/CoS
  - Widespread end-to-end TCP/IP QoS deployment most likely by middle to late PM-1 era
- ATM QoS
  - ATM QoS *could* be used end-to-end *today* over ATM WANs (NREN/NISN, ESnet, vBNS, etc.) if campus backbone is ATM and uses either ATM-to-the-server/desktop or Ethernet edge-switches (IEEE 802.1p/q etherswitches with ATM uplinks to ATM backbone)
  - ATM technology has had sufficient QoS in hardware from inception - designed as multimedia multiservice transport
  - Several classes of service defined within ATM to support QoS
- More R&D needs to be done to link TCP/IP QoS with ATM QoS and vice versa.

NASA Ames Research Center
EOSDIS Technology Assessment Team 3/30/98

20

TCP/IP Networks QoS Measurements

Frame Loss Rate: Percentage of frames that should have been forwarded by a network device under steady state (constant) load that were not forwarded due to lack of resources. FLR is defined as the percentage of packets forwarded by the source that were not returned by the device under test within an expected window of time. Frame loss is generally due to buffers overflowing because packet processing does not empty buffers as fast as they are being filled. In addition, packets that are delayed beyond the measurement window will be considered as "lost" by the tester in spite of the fact that they may not have been dropped. (FLR is usually measured by a different procedure than the one used for throughput: for each packet size and number of streams the offered load is stepped down from 100% of the maximum theoretical rate in steps of 10%. The percentage of forwarded packets that is lost by the network device is recorded until the number of packets lost is zero for two successive levels of offered load).

Latency:

For store and forward devices: The time interval starting when the last bit of the input frame reaches the input port and ending when the first bit of the output frame is seen on the output port.

For bit forwarding devices: The time interval starting when the end of the first bit of the input frame reaches the input port and ending when the start of the first bit of the output frame is seen on the output port.

Throughput: The maximum rate at which none of the offered frames are dropped by the device. This maximum rate is determined by a binary search procedure that starts at 50% of the theoretical rate and increments and/or decrements forwarding rate until the maximum lossless rate is found. The standard test suites measure unidirectional traffic for the maximum number of streams the network device can support.

ATM Networks QoS Measurements

The number and types of performance measurements of ATM systems are considerably more varied than in TCP/IP frame or packet based systems. One reason is that ATM networks were designed from the outset to support extensive QoS and traffic management parameters per virtual circuit, rather than simple best-efforts and binary-rate traffic management schemes historically available with TCP/IP. Another reason is that there are a number of traffic types or service categories implemented in ATM. For example, ATM service categories include: Constant Bit Rate (CBR), Unspecified Bit Rate (UBR) and variant UBR+, real-time Variable Bit Rate (rt-VBR), non-real-time Variable Bit Rate (nrt-VBR), Available Bit Rate (ABR), and Unspecified Bit Rate (UBR). Each service categories have pre-defined assigned Quality of Service (QoS) parameters and traffic management descriptors that establish the rules, limits and tolerances for cell flows. At this juncture, subject to further study, the EOSDIS time-critical file transfers should probably use nrt-VBR service, whether unicast or reliable multicast in nature. This will probably be optimal in establishing the bandwidth allocations required to both meet a data product distribution schedule between processing sites (by specifying the Minimum Cell Rate or MCR – see below) and best avoid the ATM equivalent of "congestion" which is essential to the smooth operation of the reliable multicast protocol layer.

There is ongoing work in the ATM Forum Testing and Traffic Management Working Groups on defining and refining the methodology of performance measurements and establishing which are meaningful and which are not. This work has been evolving and has been revised and refined since late 1994. Raj Jain of Ohio State University and his group have been the primary contributors to this body of work. A preliminary review of their extensive contributions strongly suggests adopting their measurement definitions and techniques and consulting them in engineering the NREN/NGI network to meet the EOSDIS requirements.

Reliable Multicast Protocols

- Transmit a single copy of files to arbitrarily large number of receivers *simultaneously* rather than serially and repetitively
- Significant savings of: internal network bandwidth, elapsed time to complete transmissions to all receivers, I/O and CPU at the server
- Publish/Subscribe push model where all sites requiring a data product (of any level) are subscribed and simultaneously receive them as soon as they are available through reliable multicast
- Most applicable in WAN and in the most highly distributed architectures (be they for data product production/user delivery)
- Extremely applicable when overall EOSDIS system architecture is distributed and *optimized* to use precursor standard data products (push) to *fulfill large share of ad hoc queries (pull)*
- *RM protocols still being developed and researched. Strongly recommend prototyping effort and support for work in this area.*

NASA Ames Research Center
EOSDIS Technology Assessment Team 3/30/98

21

Reliable IP and ATM Multicast protocols/technology provides a technique for simultaneous distribution of data products to a wide range of numbers of sites (e.g. tens, hundreds, thousands). Multicast protocols are employed for several reasons, depending on the specific applications. Point-to-point transfers are called unicasts and might be used to send a data product to one receiver/user at a time, either starting the next transmission after the last has completed (worst-case) or sending several at the same time to the extent that the sending server's disk I/O cycles, the network bandwidth of the server, and the supporting compute cycles in the server permit. Sending N-unicasts will generally use N-times as many resources at the source and also importantly in the Wide Area Network (WAN) where bandwidth is relatively expensive and/or constrained.

The technique of multicasting a file or a multimedia stream simultaneously to N receivers/users is that a path/route through the network is constructed from senders to receivers that is similar to the "Spanning Tree". As packets (or cells) move through the network toward their destinations, they are replicated or copied at the last possible branching point. This minimizes the traffic load inside the network. In the case of a shared network service such as NREN's use of the Sprint ATM "public cloud", the economy may go to the service provider unless reductions for multicasts are specifically negotiated. In any event, the shared network infrastructure will be less congested, allowing more calls to be set up inside of it – a direct benefit to the customer. In the case of the NASA Internet (NI), for example, where NASA acquires its own leased lines, multicast bandwidth reduction results in a direct cost reduction to NASA. Indeed, NASA's globally leading use of the "Multicast Backbone" or MBone takes advantage of best effort (unreliable, no QoS) delivery of multimedia streams like the Space Shuttle missions.

Compared to multiple unicasts, in the best case (assuming no loss), the key parameters that lead to bandwidth savings as seen on typical MBone applications are: a) each packet only traverses each link once (whereas with multiple unicasts, it may have to traverse it N (the number of receivers) times; b) most distributions are going to be in a tree shape with typical fanout of about 3-4, leaves may have about 10+ sites and a much larger number of end-node receivers. The amount of eliminated redundancy in the tree gets higher the closer you are to the source. Savings are heavily dependent on the topology of the source, network routes, and location of receivers. For example, retail giant Toys R Us uses StarBurst Multicast File Transfer Protocol (MFTP) to update software on servers at 250 U.S.-based stores. Previously, updates were unicast to each store individually, a process that took three or four days. With MFTP the same update takes about an hour.

Reliable multicast protocols are transport protocols layered on top of multicast routing protocols and add reliable transport to multiple receivers to the best-efforts delivery of IP and ATM in much the same way as TCP (the most common transport protocol) does for unicasts. They are the subject of intense research and development because they enable the next generation of revenue creating network services and relieve network congestion. They are also quite complex and vary by application, scale, etc. We have been collecting papers on a variety of reliable multicast protocols and on-going research projects. In anticipation of the use of reliable multicast services by EOSDIS and perhaps NASCOM, and because this is clearly a key new technology and service for the Next Generation Internet (NGI), we are in the process of initiating benchmarking of several protocols and establishing Joint Sponsored Research Agreements with leaders in academia and industry. More precise estimates of bandwidth savings must await more specific topology information of the alternative EOSDIS architectures being studied and the typical scales of data product publish/push and file sizes.

Most of the currently proposed and implemented reliable multicast protocols are based on IP. Several are available for testing and benchmarking. The Internet Research Working Group (IRWG) which performs the precursor research for the Internet Engineering Task Force (IETF) who establish Internet standards, is now working on several reliable multicast protocols. Extensions to ATM for reliable multicast are being researched and proposed by groups at Washington University, by Jon Turner, and elsewhere.

Shared Frame Relay

- Cost effective alternative for WAN connectivity of T3 and below
- Widely available and mature service
- Interoperable with both legacy technologies and newer technologies such as ATM

- NISN is deploying Frame Relay connected sites together with the NISN ATM service for AEROnet using Sprint

NASA Ames Research Center
EOSDIS Technology Assessment Team 3/30/98

22

Frame relay describes an interface standard. As such, it is a technology that has been optimized for the transport of protocol-oriented data in discrete units of information (generic packets). Its ability to statistically multiplex provides the same bandwidth-sharing and efficiency as X.25. With traditional time division multiplexing and circuit switching, each path (circuit) through the network has dedicated bandwidth allocated to that particular circuit on a static basis for the duration of the call. For example, in a traditional voice call, just as much bandwidth is used to transmit the silence (when listening and pausing between words and sentences) as the sounds when actually speaking.

Virtually all data has similar "silence" between transmissions. In fact, data transmissions are typically much more "bursty" than voice conversations. Thus, when dedicated bandwidth is used, the transmission facilities are essentially unused a large percentage of the time.

By contrast, statistical multiplexing means that paths (virtual circuits) are defined through the network. However, no bandwidth is allocated to the paths until actual data (real information) needs to be transmitted. Then, the bandwidth within the network is dynamically allocated on a packet-by-packet basis. If, for a short period of time, more data needs to be transmitted than the transmission facilities can accommodate, the switches within the network will buffer (store) the data for later transmission. In the event that this over subscription persists, congestion control mechanisms must be invoked. Frame relay also eliminates much of the protocol processing done by the network, thereby reducing the portion of the transmission latency attributed to protocol processing. The simplification of the protocol focuses on the elimination of error recovery functions. Instead, the element of the system that guarantees the error-free end-to-end transfer of frames are the endpoint devices (e.g. PCs or workstations), not the network itself. This protocol processing, which is still necessary to guarantee the accurate delivery of the data, is left to the higher layers inherent in the transported data.

The frame relay interface specification provides a signaling and data transfer mechanism between endpoints and the network. This interface allows communication bandwidth to be shared among multiple users, creating instantaneous bandwidth allocation on demand. Each frame (or packet) contains header information that is used to determine the routing of the data to the desired destination. This enables each endpoint to communicate with multiple destinations via a single access link to the network. Instead of fixed amounts of bandwidth allocated to the resource, frame relay traffic receives full bandwidth for short transaction bursts.

A frame relay network consists of user devices and network devices that implement the standard interface. The user device is responsible for delivering frames to the network in the prescribed format. The network is responsible for switching or routing the frames to the proper destination user device. Benefits of Frame Relay include: reduced internetworking costs, increased performance with reduced network complexity, and increased interoperability via international standards.

Reduced Internetworking Costs. When using a private frame relay network, statistically multiplexed traffic from multiple sources over private backbone networks can reduce the number of circuits and corresponding cost of bandwidth in the wide area. Public frame relay services almost universally save money when compared with the equivalent service provided by dedicated leased lines.

Since frame relay provides multiple logical connections within a single physical connection, access costs are also reduced. Equipment costs may be lowered by reducing the number of port connections required to access the network. For remote access devices, access line charges can be lowered by reducing the number of physical circuits needed to reach the networks.

Increased Performance with Reduced Network Complexity. Both by reducing the amount of processing (as compared to X.25) and by efficiently utilizing high speed digital transmission lines, frame relay can improve performance and response times of applications.

Frame relay reduces the complexity of the physical network without disrupting higher level network functions. In fact, as discussed earlier, it actually utilizes the existence of these higher layer protocols to its advantage. It provides a common network transport for multiple traffic types while maintaining transparency to higher level protocols unique to the individual traffic types. The frames contain addressing information that enables the network to route them to the proper destination.

Increased Interoperability via International Standards. Frame relay's simplified link layer protocol can be implemented over existing technology. Access devices often require only software changes or simple hardware modifications to support the interface standard. Existing packet switching equipment and T1/E1 multiplexers often can be upgraded to support frame relay over existing backbone networks.

Frame relay is an accepted interface standard that vendors and service providers are adhering to and implementing. Most areas of the frame relay standards are well-defined and have been approved by ANSI and the ITU/TSS (formerly known as CCITT). The simplicity of the frame relay protocol accommodates quick and easy interoperability testing procedures between devices from different vendors, and this interoperability testing is currently in progress among vendors as are certification processes for carriers providing frame relay services.

Private/Dedicated Networks

- Private networks must be conservatively engineered for link failures and over engineered to handle loads
- As a result, they are Costly
- No flexibility in the sizing of dedicated circuit bandwidths.
- Very mature technology and almost ubiquitous.
- Private networks require back-hauling to closest backbone site

- A business case was presented to NISN which achieved a 56% cost savings by moving the NISN backbone from a dedicated/private network to a shared ATM network[1]
  - EBnet and external networks should be considered for migration to ATM as well

1 - "ATM Proposal for NISN and NREN/HPCC", 9/96 Blaylock and Lisotta

NASA Ames Research Center
EOSDIS Technology Assessment Team 3/30/98 23

T- Carrier is the generic name for any of several digitally multiplexed carrier systems. North American designators for T carriers correspond to the digital service (DS) level hierarchy shown in table 1 below.

Table 1: Digital Service (DS) Level Hierarchy

DS Level	Speed
DS0	64Kbps
DS1	1.544Mbps
DS1C	3.152 Mbps
DS2	6.312 Mbps
DS3	44.736 Mbps
DS3C	89.472 Mbps
DS4	274.176 Mbps

The T-carrier was the first type of data transfer mechanism used in the early days of the Internet and is a private line dedicated between two points. The bandwidth of a T-carrier line used in data communications is typically dedicated per each network connection. Therefore in certain environments there is inefficient use of the available bandwidth because when the node connected to the line is not using the circuit, the circuit is idle. The tradeoff is that there is little delay and this technology has been proven for many years and is provisioned cost-effectively by carriers over ordinary copper lines in the metropolitan area. Typical EOSDIS site connections might be made over one or a few T-1 or DS1 circuits or a single DS-3 or T-3 circuits. One important trend is that the relative cost of T-3 to T-1 circuits is dropping so that the crossover between multiple T-1s and a T-3 is now in the 4-6:1 range while the bandwidth is 45:1.5 or around 30:1. In general, there are improving economies of scale in buying higher bandwidth circuits.

LEO Broadband Data Satellites

- Global fiberoptic-like connectivity w/constellations of many small satellites capable of on-board switching and intersatellite linking
  - designed for broadband data; not voice b/w (e.g. Iridium)
  - VSAT earth station w/DBS-sized antenna (~18”) & PC-sized cost
  - 72msec round-trip propagation delay comparable to Internet’s
  - Bit error-rates (BERs) in the 10exp-6 range or better
  - available *anywhere* on earth, not just in major metro areas
- Satellites are inherently a broadcast/multicast medium
  - LEO inter-satellite transmissions provide multicast routing
  - footprints of multicast route member satellites provide regional to global coverage, as required by a given data product distribution
  - hybrid terrestrial low-bandwidth return paths when available
- Players: Teledesic Corp. (2002 operation), Motorola Celestri, others?
- NASA LeRC is COE for satellite communications

NASA Ames Research Center
EOSDIS Technology Assessment Team 3/30/98 24

The use of geosynchronous satellites for transmission of video, voice, and data has existed for over two decades. One of the most significant benefits of satellite-based transmission arises from the “footprint” its antenna has on the earth. For domestic satellites, this includes the Continental U. S. and spot-beams for Alaska and Hawaii, for example. This makes satellites an extremely powerful way of broadcasting the same information to an arbitrarily large number of receivers in the footprint. Witness the recent assault on the CATV marketplace by the direct broadcast video satellite (DBS) vendors using receivers that sell for as little as $100. The stationary or geosynchronous orbits around the earth’s equator require the satellites to be positioned approximately 22,500 miles above the earth. At the speed of light this represents a round-trip propagation delay of approximately 500milliseconds for single-hop data transmissions. Obtaining high throughput at high data rates requires tuning and even extension of the protocols used. Still, networks like the NASA Internet (NI) and the Internet use satellite links to economically connect countries and areas of the earth that are under-served by terrestrial connectivity options.

In the 2001-2002 time-frame a totally new generation of satellites that are specifically designed for global broadband data (rather than voice or video) services will be operational. Currently there are at least three Low-Earth Orbit (LEO) satellite systems in development: Teledesic Corporation’s (initially backed by Bill Gates of Microsoft and wireless telecommunications pioneer Craig McCaw), Motorola's Celestri (a response to losing the Teledesic construction contract to Boeing) and M-Star systems. The LEO orbits will be approximately 320 miles above the earth and offer round-trip propagation times of only 72milliseconds. These delays are comparable to best-case terrestrial cross-country delays on the Internet. These shorter delays combined with the ability to engineer bit-error-rates in the range of 10exp-6, allows the vendors to claim “fiberoptic-like” connectivity.

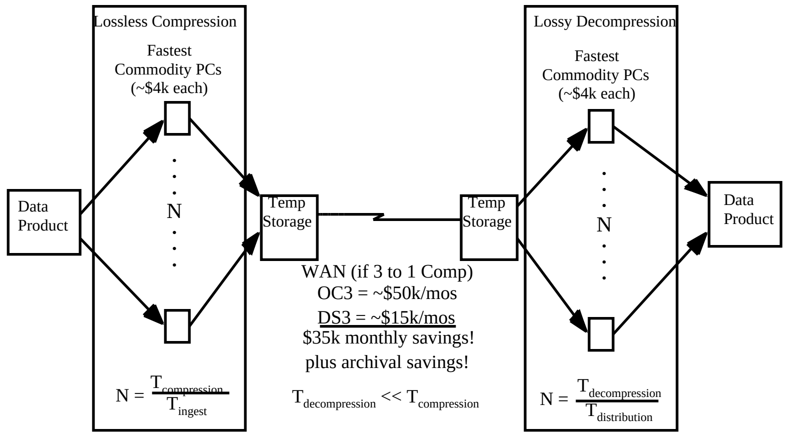
Using a constellation of up to several hundred satellites, they will provide a global, broadband ATM "Internet-in-the-sky". Applicability to EOSDIS, subject to the ultimate pricing of services, is very likely to include providing access for users in under-served regions of the world (no local/metro-area fiberoptic connectivity) and very possibly as a broadcast medium in a hybrid (LEO-broadband forward path, Internet-NACK-return-path) reliable multicast environment for distributing highly subscribed-to data products. The earth station for this 30GHz uplink/20GHz downlink system will be similar to DBS antenna sizes and PC workstation costs. ATM connectivity and bandwidth assignment will be on-demand switched virtual circuits (SVCs) of from 16Kbps to over 2Mbps in 16Kbps steps and at asymmetric rates. Teledesic’s system will be able to accommodate special uplinks of up to OC-3 or 155Mbps. In its distributed architecture, dynamic routing, and robust scalability, the Teledesic Network emulates the most well known distributed network, the Internet, while adding the benefits of real-time capability and location-insensitive access.

Clearly satellite bandwidth should cost more than terrestrial bandwidth which is based on the ultimate capacity of fiber pairs to carry terabits/second of information. The applicability of LEO satellites to EOSDIS for use in query-response or “pull” scenarios, will therefore most likely be limited to users without less expensive connection options like the Internet if it can offer differentiated non-congested services. However, if EOSDIS does make significant use of the “publish-subscribe” or “push” multicast protocols, some economies of bandwidth use may allow additional use to “prove-in” economically. This would be a function of the number of subscribed receivers for standard information products, their geographic distribution (LEO satellite-to-satellite transmission is essentially the same as LAN/WAN multicast routing trees), and how the vendors charge for multicast distribution. Unlike network access based on leased lines which have a flat-rate monthly fee, LEO costs are likely to be dominated by usage. The ultimate applicability will become clearer as LEO unicast and multicast bandwidth pricing is established as well as those for the emerging Internet access options (xDSL, CATV, WATM and other wideband terrestrial wireless services) are widely available and established in the next few years.

Compression

- Seems to a natural fit to the EOSDIS application.
- If used, significant savings can be achieved from the WAN due to the reduction in the amount of capacity that will be needed.
- A "back of envelope" analysis using fast commodity PC to handle the compression of MODIS data products revealed a monthly savings of $35K on a single WAN access link (assuming going from an OC3 to DS3). The assumption was a 3 to 1 compression of the data. The ROI proved to be one to two months.
- Note that this analysis was a mathematical exercise but it revealed that compression could result in significant cost savings. We are recommending that EOS invest in a prototype to determine if compression can be used and how it would be used in the EOS system.

Compression



For example MODIS:

T_{ingest}	= 5 min	T_{ingest}	$= T_{distribution}$
$T_{compression}$	= 20 min to compress 1 GB file	$T_{decompression}$	= 5 min
N	= 20/5 = 4 PCs at $4k each = $16k	N	= 5/5 or $4k

Note: Numbers presented here have not been verified and are educated guesses to be used for discussion only.

NASA Ames Research Center
EOSDIS Technology Assessment Team 3/30/98

26

WAN Technologies Summary

Architecture Independent	Shared ATM			Shared Frame Rly			Satellite LEOs			Network QoS/CoS			Reliable Multicast			Compress-ion			Dedicated Circuits		
Options 1-4 including user access	A M	P M	Ch m	A M	P M	Ch m	A M	P M	Ch m	A M	P M	Ch m	A M	P M	Ch m	A M	P M	Ch m	A M	P M	Ch m
Maturity	G	G	G	G	G	G			R	R	Y	G	R	Y	G	G	G	G	G	G	G
Scalability (up/down)	Y	G	G	Y	Y	Y			R	Y	G	G	Y	G	G	G	G	G	G	G	G
Functionality	Y	G	G	G	G	G			R	G	G	G	G	G	G	Y	G	G	G	G	G
Cost	G	G	G	G	G	G			R	G	G	G	G	G	G	G	G	G	R	R	R
Interop (w/legacy & installed base)	Y	G	G	G	G	G			G	R	Y	G	R	G	G	G	G	G	G	G	G
Performance (Capacity, reliability, delay)	G	G	G	G	G	G			G	Y	G	G	Y	G	G	G	G	G	G	G	G
Market Trend	G	G	G	G	G	G			G	G	G	G	G	G	G	G	G	G	Y	R	R
Mature Standards	G	G	G	G	G	G			R	R	Y	G	R	Y	G	G	G	G	G	G	G
Availability	Y	Y	G	G	G	G			G	R	G	G	R	Y	G	G	G	G	G	G	G
Recomend	G	G	G	G	G	G			R		G	G	R	G	G	Y	G	G	Y	Y	Y

Y - IP QOS
G - ATM QOS

Key: Blank = Not Applicable R = applicable but not ready, prototype/invest Y = limited deployment G = fully deploy